ARTICLES

Effect of Estrogen Replacement Therapy on the Specificity and Sensitivity of Screening Mammography

*Mary B. Laya, Eric B. Larson, Stephen H. Taplin, Emily White**

Background: **Previous studies have demonstrated that mammographic breast density increases following the initiation of estrogen replacement therapy (ERT). The effect, if any, that this increase in density has on the specificity (related to false-positive readings) and the sensitivity (related to false-negative readings) of screening mammography is unknown.** ***Purpose:*** **Using a retrospective cohort study design, we assessed the effects of ERT on the specificity and the sensitivity of screening mammography.** ***Methods:*** **Participants (n = 8779) were postmenopausal women, aged 50 years or older, who were enrolled in a health maintenance organization located in western Washington state and who entered a breast cancer screening program between January 1988 and June 1993. Two-view mammography was performed as part of a comprehensive breast cancer screening visit. Menopausal status, as well as demographic and risk-factor information, was recorded via self-administered questionnaires. Hormonal replacement therapy type and use were determined from questionnaire data and from an automated review of pharmacy records. Individuals diagnosed with breast cancer within 12 months of their first screening-program mammograms were identified through use of a regional cancer registry. Risk ratios (RRs) plus 95% confidence intervals (CIs) of false-positive as well as false-negative examinations among current and former ERT users (with never users as the reference group) were calculated. Reported *P* values are two-sided.** ***Results:*** **The specificity of mammographic screening was lower for current users of ERT than for never users or former users. Defining a positive mammographic reading as any non-normal reading (either suspicious for cancer or indeterminate), the adjusted RR (95% CI) of a false-positive reading for current users versus never users was 1.33 (1.15-1.54) (*P*<.001); for former users versus never users, the RR (95% CI) was 1.00 (0.87-1.15). The adjusted mammographic specificities (95% CIs) for never users, former users, and current users of ERT were 86% (84%-88%), 86% (84%-87%), and 82% (80%-84%), respectively. Defining a positive reading more rigorously (i.e., as suspicious for cancer only), the adjusted RRs (95% CIs) of false-positive readings for current users and former users (versus never users) were 1.71 (1.37-2.14) (*P*<.001) and 1.16 (0.93-1.45), respectively. Sensitivity was also lower in women currently receiving ERT. The unadjusted RR (95% CI) of a false-negative reading for current users versus never users was 5.23 (1.09-25.02) (*P* = .04); for former users versus never users, the RR (95% CI) was 1.06 (0.10-10.87). The unadjusted mammographic sensitivities (95% CI) for never users, former users, and current users of ERT were 94% (80%-99%), 94% (69%-99%), and 69% (38%-91%), respectively.** ***Conclusions and Implications:*** **Current use of ERT is associated with lower specificity and lower sensitivity of screening mammography. Lower specificity could increase the cost of breast cancer screening, and lower sensitivity may decrease its effectiveness. [J Natl Cancer Inst 1996;88:643-9]**

Screening mammography is one of the most important and effective means of early detection of breast cancer, and its use has been associated with a reduction in breast cancer mortality among women 50 years of age and older (*1*). As with any screening method, the performance characteristics—the sensitivity and the specificity—play a large role in determining the effectiveness of early detection programs (*1-8*). Postmenopausal estrogen replacement therapy (ERT) may influence the performance characteristics of screening mammography through its effect on radiographic breast density.

The degree of mammographic breast density changes throughout a woman's life (*9-14*), and, in general, the premenopausal breast is radiographically dense. Marked radiographic breast density is associated with a failure to diagnose cancer by mammography (*15-17*). This fact is thought to ac-

**Affiliations of authors:* M. B. Laya, E. B. Larson (Department of Medicine), S. H. Taplin (Department of Family Medicine), E. White (Department of Epidemiology), University of Washington, Seattle.

Correspondence to: Mary B. Laya, M.D., M.P.H., Department of Medicine, University of Washington, 4245 Roosevelt Way, NE, Seattle, WA 98105-6920.

See "Notes" section following "References."

count, in part, for the decreased sensitivity of mammographic screening among younger women (*18-20*). Specificity may also be affected, since the radiologist's degree of certainty is reduced and the number of false-positive examinations increases in association with marked density (*9*). With age, and particularly following menopause, glandular elements within the breast regress and are replaced by fat, resulting in increased radiolucency and better performance characteristics for screening mammography.

This normal involutional process may be retarded or reversed by the use of postmenopausal ERT. Hormonal replacement therapy results in an increase in mammographic breast density or in a change in the parenchymal pattern in a significant proportion of postmenopausal women (17%-73%, depending on how the changes are measured) (*12,21-25*). An ERT-induced increase in mammographic density might result, therefore, in an increase in false-positive and/or false-negative readings among postmenopausal women. Given that the population of women for whom screening mammography is most effective (those 50 years of age and older) is also the group most likely to be considered for ERT, any deterioration in the accuracy of this test associated with ERT would have potentially important implications.

The aim of our study was to determine the effects of ERT on the performance characteristics of screening mammography. We sought to define the frequencies of false-positive and false-negative examinations to compare the specificities and the sensitivities of mammography for women who are current or former users of ERT and for those who have never used ERT. Our study population included women enrolled in the Group Health Cooperative (GHC) of Puget Sound, WA, and we used a retrospective cohort study design.

Subjects and Methods

Subjects were enrollees in the GHC, a closed-panel health maintenance organization with 400 000 members that includes two hospitals and 21 primary care facilities in western Washington state. Female enrollees were slightly better educated than the general population of comparably aged women and were more likely to have household incomes above $15 000 and below $35 000 than the general population of western Washington. A Breast Cancer Screening Program (BCSP) was implemented in 1985 (*26,27*) to achieve regular and appropriate use of mammography among GHC members. All female enrollees 40 years of age or older were mailed a two-page questionnaire on enrollment or shortly after their 40th birthday. At the time of this study, 84% of the GHC women enrollees over 40 years of age had completed the questionnaire, which elicited information concerning breast cancer risk factors, reproductive history, screening history, and past estrogen use (e.g., "Have you ever taken estrogen or other hormones for menopausal symptoms or prevention of osteoporosis?") (*27*). All women 50 years of age and older and selected women 40-49 years of age who were at high risk of developing breast cancer were invited to undergo screening at 1-2-year intervals. The BCSP screening visit included a clinical breast examination, instruction in breast self-examination, and two-view mammography. Screened women were asymptomatic; those with breast complaints or self-discovered lumps were referred to their primary care provider for evaluation and did not enter the screening program. Radiographers were unaware of the results of the breast physical examination, which was performed by the screening center nurse-clinician, and they did not have access to responses recorded on the BCSP screening questionnaire. The radiology department did request that patients complete a brief questionnaire regarding past and current breast symptoms, family history of breast cancer, and current hormone use for inclusion in the x-ray film jacket. Thus, the radiographer had access to some clinical information concerning those screened.

The three screening centers with mammographic facilities that provided data for this study were evaluated and approved by the American College of Radiology. The project was approved by the Institutional Review Board of Group Health Cooperative of Puget Sound and by the University of Washington.

The BCSP data are maintained on an IBM mainframe computer that contains all survey data as well as screening visit and mammographic findings and recommendations. On a quarterly basis, the Seattle–Puget Sound Surveillance, Epidemiology, and End Results (SEER)[1] cancer registry provides the GHC with an extract from its database detailing all cases of cancer that are marked as diagnosed and/or treated among women enrolled in the GHC. Internal SEER audits demonstrate nearly 100% case ascertainment for cancers diagnosed among GHC women (*28*). The GHC computerized pharmacy database was used to retrospectively assess hormone use during the 12 months prior to BCSP mammogram completion. Each pharmacy record includes information about the drug prescribed and its prescription date as well as information about its dose, the quantity dispensed, and the dosing instructions.

Identification and Exposure Classification of Subjects

Subjects were selected from women enrolled in the BCSP on the basis of the following criteria: 1) age of 50 years or older with a survey response indicating natural cessation of menses or age of 55 years or older with a survey response indicating surgical menopause, 2) continuous enrollment in the GHC for at least 12 months before and 12 months after the mammographic screening date, and 3) first BCSP screening visit between January 1988 and June 1993. Only x-ray films from a woman's first BCSP mammography visit were included in the analysis; these mammographic films may not be a subject's first screening films, although a history of prior non-BCSP mammography is included in the analysis. Subjects were excluded from the study if they had a personal history of breast cancer before entry or they failed to respond to the BCSP survey question concerning past estrogen use.

Hormone use for each subject was classified on the basis of survey response data and on pharmacy records covering the period from the date of survey questionnaire completion to the date of the first screening visit. A never user was defined as an individual with survey-response data indicating no past ERT use and no prescription for estrogen recorded in the pharmacy database from the date of survey completion to the date of the screening visit. Current user status was determined by searching the pharmacy database for the last hormone prescription preceding the screening visit date. A woman who took at least 80% of the prescribed dose, based on refill information, and who received enough pills to last until the screening visit date was counted as a current user. Among current users, 94% had taken ERT for 3 months or more during the year preceding the mammographic screening date. A former user was defined as a noncurrent user who had indicated use of estrogen on the BCSP survey questionnaire or who had at least one prescription for estrogen during the interval between completing the survey and the screening visit but who was not taking ERT at the time of mammography. A similar scheme was used to classify progesterone use.

Classification of Mammographic Readings and Outcomes

According to the BCSP program, mammograms were classified as "negative" (no radiographic evidence of cancer), "indeterminate" (mammogram with features requiring further evaluation), or as "possessing features suggestive of malignancy." In general, a positive screening mammogram can be defined in a number of ways, and this definition affects the specificity and sensitivity calculations (*1*). For our primary analysis, we defined a positive mammogram as any non-normal reading, including both indeterminate readings and those suggestive of cancer. This liberal definition of a positive test would maximize the detection of all false positives, even those leading to a low-level follow-up recommendation (e.g., repeat mammography at 6 months). This definition of a positive mammogram (suggestive of cancer plus any abnormal) was also used to calculate the sensitivity of the screening test. Specificity calculations were further performed using a second definition that included as positive only those mammograms that were read as suggestive of cancer.

SEER data were used to determine whether a diagnosis of breast cancer, including ductal carcinoma in situ (DCIS), occurred within 12 months of the screening visit. Cases of lobular carcinoma in situ were not included. The SEER staging system was used to classify the extent of disease (stage 0 = carcinoma in situ, stage 1 = local disease, stage 2 = regional disease, and stage 3 = distant spread of disease).

A negative reading was designated a true negative if no breast cancer was diagnosed within 12 months of the screening date for mammograms originally classified as no signs of cancer. A reading was defined as a false negative if

breast cancer was diagnosed within 12 months of a negative screening mammogram (an interval cancer). A positive reading was classified as a true positive if breast cancer was diagnosed within 1 year of the positive reading or a false-positive examination if no cancer was diagnosed within 1 year.

Specificity and the false-positive frequency (1 – specificity) are complementary ways of relating how often a test appropriately classifies normal subjects without disease. Specificity was defined as the proportion of negative mammograms among women without cancer. The false-positive frequency is the proportion of those subjects without cancer but with a positive examination. Sensitivity and the false-negative frequency (1 – sensitivity) relate how well a test identifies those subjects with disease (*29*). Sensitivity was defined as the proportion of cancer cases with an associated abnormal mammogram. The false-negative frequency is the proportion of cancer cases with an associated negative examination.

Statistical Analysis

In comparing exposure group characteristics, we used the Student's *t* test for continuous variables and the chi-squared test for categorical variables. All statistical tests were two-tailed. Comparisons of and confidence intervals (CIs) for binomial proportions were computed using the normal theory method. When the sample size was small, we employed Fisher's exact test and determined the CIs using the exact method. Separate analyses were conducted for false-positive mammograms among women who did not have breast cancer (specificity analysis) and for false-negative mammograms among women who did (sensitivity analysis). Stratification and logistic regression methods were used to control for potential confounders of the relationship between ERT exposure group and the outcome of mammographic screening using unconditional logistic regression (*30*). Potential confounders were chosen based on the published literature and included factors known to affect the sensitivity and specificity of mammography (*8*) as well as factors that might affect breast density (*11,13,31-34*). Variables evaluated as potential confounders included age, history of a first-degree relative with breast cancer, existence of a previous mammogram, age at menarche of less than 11 years, weight, height, body mass index, age greater than 29 years at first full-term pregnancy, year of index mammogram, current smoking, and past oral contraceptive use for 1 year or more. Adjusted specificities were derived from the regression equation using mean values for the final adjustment variables for the entire cohort.

Results

Study Population Characteristics and Outcomes

The characteristics of the study population and of the three ERT exposure groups are shown in Table 1. The groups exhibited statistically significant differences. Current users of ERT were younger, had a lower body mass index (weight in kilograms divided by height in meters squared), were more likely to have had a previous mammogram and a previous breast biopsy, and were more likely to be surgically menopausal than never users. Current users also had a lower prevalence of nulliparity, late age at first full-term birth, and early menarche. Conjugated equine estrogen (Premarin, Wyeth-Ayerst, Philadelphia, PA) was the estrogen used by 80% of current users, and esterified estrogen was used by 19%. Virtually all of the progesterone received was medroxyprogesterone acetate.

Table 2 summarizes the outcomes of the first BCSP screening mammograms for the three groups. The overall cancer prevalence was 7.2 per 1000. Cancer prevalence was slightly higher among those who had never used ERT, a difference that was not statistically significant ($P = .14$ for current users versus never users). Sixty-three cancers were detected in the study group at 1 year; 56 were associated with an abnormal mammogram (true positives), and seven were associated with a normal mammogram (false negatives). Forty-five (71%) of the cancers were early stage breast cancers, i.e., DCIS (n = 2) or stage I (n = 43). Thirty-six percent of all tumors measured 15 mm or less in size, and 23% were 10 mm or less. The mean size of the tumors did not differ between the three ERT exposure groups.

Specificity

The unadjusted specificity of screening mammography (with a positive defined as any abnormal) was lower for current users of ERT (82%) than for former users or never users (85%) (Table 2). Because of the presence of significant differences in the clinical characteristics of the three ERT groups, we employed multiple logistic regression to adjust for confounders of the relationship between the false-positive frequency (1 – specificity) and ERT use among women without cancer. The risk ratios (RRs) listed in Table 3 reflect the proportion of false-positive examinations among women without cancer in former and

Table 1. Mean values of selected characteristics of study subjects according to estrogen replacement therapy (ERT) exposure group

Variable	All subjects (n = 8779)	Never used ERT (n = 3826)	Former user of ERT (n = 2853)	Current user of ERT (n = 2100)	*P**
Age at first BCSP† mammogram, y	65.8	66.5	66.9	63.1	<.001‡
Weight, kg	68	69	68	67	<.001‡
Body mass index, kg/m^2	25.9	26.3	25.9	25.1	<.05‡
Surgical menopause, %	30.4	15.6	36.5	48.9	<.001§
First-degree relative with breast cancer, %	9.4	9.5	9.5	8.8	NS§
Nulliparous, %	8.8	9.9	7.9	7.4	.002§
Age at first delivery ≥30 y, %	7.5	9.2	6.9	5.0	<.001§
Menarche before age 12 y, %	4.2	4.4	4.3	3.9	NS§
Previous breast biopsy, %	11.8	10.6	12.0	14	<.001§
Previous non-BCSP† mammogram, %	64	57.7	65.9	74.0	<.001§
Current smokers, %	16.2	16.4	16.7	15.0	NS§
History of oral contraceptive use for 1 year or more, %	22	20.4	19	31.1	<.001§

*All *P* values (two-sided) for current users versus never users. NS = not significant.
†BCSP = Breast Cancer Screening Program.
‡Student's *t* test.
§Chi-squared test.

Table 2. Mammography outcomes with 1 year of follow-up in never users, former users, and current users of estrogen replacement therapy*

	Never used, n = 3826	Former users, n = 2853	Current users, n = 2100
True negatives	3228	2427	1720
False positives	564	410	367
True positives	32	15	9
False negatives	2	1	4
Cancer prevalance per 1000	8.8	5.6	6.2†
Specificity (95% confidence interval)	85% (84%-86%)	85% (84%-87%)‡	82% (81%-84%)§
Sensitivity (95% confidence interval)	94% (80%-99%)	94% (69%-99%)‡	69% (38%-91%)‖

*Positive is defined as a reading of suspicious for cancer or indeterminate. Negative is defined as no evidence of cancer.
†Not statistically significant from former users or never users at the .05 level (normal theory method).
‡Former users and never users not significantly different at the .05 level for sensitivity or specificity calculations.
§Two-sided $P = .006$ for current users versus never users and two-sided $P = .001$ for current users versus former users (normal theory method).
‖Two-sided $P = .04$ for current users versus never users and two-sided $P > .05$ for current users versus former users (exact method).

current users of ERT. Never users comprised the referent category. We adjusted for age, history of a first-degree relative with breast cancer, and history of a previous non-BCSP mammogram. None of the other potentially confounding variables significantly affected the relationship between ERT exposure and the false-positive frequency.

Table 3 shows the unadjusted and adjusted RRs for former and current users and the corresponding unadjusted and adjusted specificities. Current use of ERT was associated with an increased likelihood of a false-positive mammographic reading when compared with never use (adjusted RR = 1.33; 95% CI = 1.15-1.54; $P<.001$). The RRs for former users were not significantly different from 1.0. Adjusted specificities were derived by entering the mean values of important covariates into the logistic regression model and were 86%, 86%, and 82% for former users, never users, and current users, respectively. Follow-up mammograms (additional views or a contracted screening interval) were ordered proportionally more often in current users (11.6%) than in former users or never users (7.0% and 7.7%, respectively; $P<.001$).

Among current ERT users, the risk of a false-positive screen was similar for those who used estrogen alone and those who used estrogen plus progesterone (Table 4). The effect of current estrogen use on the false-positive frequency was not modified by age ($P = .73$) or by the existence of a previous non-BCSP mammogram ($P = .62$). Current ERT users with a positive family history had a higher RR of a false-positive examination than those current users without a family history (Table 5), but this difference was not statistically significant ($P = .14$).

Similar specificity calculations were done after defining a positive examination more rigorously, i.e., as a reading of "suspicious for cancer" only. Unadjusted and adjusted specificities calculated in this manner are listed in Table 6 as are the RRs of a false-positive test.

Sensitivity

A higher false-negative frequency was observed among women who were current users of ERT than among never users or former users (Fisher's exact test; $P = .04$). Four (31%) of 13 cancers in the current user group were not detected by mammography, compared with two (6%) of 34 and one (6%) of 16 in the never users and former users groups, respectively. These seven false-negative cancers all measured 15 mm or more in diameter. The RR of a false-negative mammogram among women who had cancer for former users versus never users was 1.06 (95% CI = 0.10-10.87); for current users versus never users, the RR was 5.23 (95% CI = 1.09-25.02; $P = .04$). Because of the small absolute numbers of cancers, we were unable to ad-

Table 3. False-positive frequency analysis (positive defined as any abnormal reading*): the relationship of estrogen replacement therapy (ERT) to the proportion of false-positive mammograms among women who do not have breast cancer (n = 8716)

ERT use category	Unadjusted risk ratio (95% confidence interval)	Adjusted† risk ratio (95% confidence interval)	Unadjusted specificity (95% confidence interval)	Adjusted† specificity (95% confidence interval)
Never used‡	1.0	1.0	85% (84%-86%)	86% (84%-88%)
Former user	0.98 (0.85-1.12)§	1.00 (0.87-1.15)§	85% (84%-87%)§	86% (84%-87%)§
Current user	1.22 (1.06-1.40)‖	1.33 (1.15-1.54)¶	82% (81%-84%)	82% (80%-84%)¶

*Suspicious for cancer or indeterminate.
†Adjusted for age, history of first-degree relative with breast cancer, history of a previous mammogram.
‡Reference group.
§Former and never users not significantly different at the .05 level.
‖Two-sided $P<.01$ for current users versus never users and two-sided $P<.001$ for current users versus former users.
¶Two-sided $P<.001$ for current users versus never users and two-sided $P = .001$ for current users versus former users.

Table 4. False-positive frequency (positive defined as any abnormal reading*) and therapy type: the relationship of type of estrogen replacement therapy (ERT) to the proportion of false-positive mammograms among women who do not have breast cancer

ERT category	Unadjusted risk ratio (95% confidence interval)	Adjusted risk ratio† (95% confidence interval)
Not a current user (n = 3826)‡	1.0	1.0
ERT alone (n = 1265)	1.20 (1.02-1.43)	1.31 (1.10-1.55)
ERT plus progesterone (n = 835)	1.25 (1.02-1.52)	1.37 (1.11-1.68)

*Suspicious for cancer or indeterminate.
†Adjusted for age, first-degree relative with breast cancer, history of a previous mammogram.
‡Reference group.

Table 5. Family history and false-positive frequency (positive defined as any abnormal reading*): the relationship between estrogen replacement therapy (ERT) and the proportion of false-positive readings in women who do not have cancer (false-positive frequency)

ERT use category	Adjusted risk ratio† (95% confidence interval) in women with a negative family history	Adjusted risk ratio† (95% confidence interval) in women with a positive family history
Never used‡	1.0	1.0
Former user	0.97 (0.84-1.12)	1.59 (0.96-2.64)
Current user	1.28 (1.10-1.50)§	2.07 (1.2-3.5)§

*Suspicious for cancer or indeterminate.
†Adjusted for age, first-degree relative with breast cancer, history of a previous mammogram.
‡Reference group.
§Chi-squared test for interaction, positive family history versus negative family history, not significant (two-sided $P = .14$).

just for confounders. Elimination of the two DCIS cases from consideration as true positives did not change the sensitivity values. Both DCIS cases occurred in never users who had not had a previous mammogram and were detected by the first BCSP mammogram.

Discussion

We have conducted a search of the medical literature, and we believe that this is the first study to explore the relationship between ERT and the performance characteristics of screening mammography. We observed a significant and potentially important decrease in specificity among current users of ERT. This finding persisted after adjustment for variables that might independently affect the interpretive process. In addition, we found a statistically significant reduction in sensitivity of screening mammography when women currently using ERT were compared with those who were not currently receiving this therapy. Because of the small number of cancer cases, the sensitivity confidence intervals were wide, and we were unable to adjust for potentially important confounders.

The particular strengths of this study include the following: 1) its generalizability, in that it was conducted in the context of a community-based practice; 2) the existence of a comprehensive pharmacy database, allowing accurate ascertainment of current ERT use; and 3) the existence of survey information on many variables that might confound the relationship between ERT use and mammography outcome.

Our study does have several limitations. First, since we confined our investigation to asymptomatic women 50 years of age and older, the results cannot be generalized to younger women who are receiving ERT. Second, we relied on survey data to identify postmenopausal subjects and past use of ERT; thus, there is a potential for misclassification of subjects. Third, we were unable to explore the relationship between total duration of ERT use and changes in mammographic performance characteristics. An analysis of this relationship constitutes an important area for further study.

We suggest several possible explanations for the association between ERT use and an increase in the false-positive frequency and a reduction in the specificity of screening mammography. First, postmenopausal ERT results in an increase in radiographic breast density. Fajardo et al. (*9*) have demonstrated that a radiologist's certainty of interpretation of a mammogram is inversely related to the density and complexity of the image. This uncertainty may lead to a greater number of "abnormal" readings, i.e., an increase in the false-positive frequency, and the need for supplemental diagnostic tests. An increase in density might also decrease the sensitivity of mammographic screening by obscuring important details. Several studies (*16,17,35*) have

Table 6. False-positive frequency analysis (positive defined as a reading suspicious for cancer only): the relationship of estrogen replacement therapy (ERT) to the proportion of false-positive mammograms among women who do not have breast cancer (n = 8716)

ERT use category	False positives	True negatives	Unadjusted risk ratio (95% confidence interval)	Adjusted risk ratio* (95% confidence interval)	Unadjusted specificity (95% confidence interval)	Adjusted specificity* (95% confidence interval)
Never used†	187	3605	1.0	1.0	95% (94%-96%)	95% (94%-96%)
Former user	156	2681	1.12 (0.90-1.39)	1.16 (0.93-1.45)	95% (94%-96%)	94% (93%-96%)
Current user	163	1924	1.63 (1.31-2.03)‡	1.71 (1.37-2.14)§	92% (93%-94%)	91% (90%-94%)§

*Adjusted for age, history of first-degree relative with breast cancer, history of a previous mammogram.
†Reference group.
‡Two-sided $P<.002$ for current users versus never users (other intergroup comparisons not significant at the .05 level).
§Two-sided $P<.001$ for current users versus never users (other intergroup comparisons not significant at the .05 level).

found an association between high breast density and a failure to detect cancer at screening.

Another explanation for the reduction in specificity might be that mammographers are more cautious in interpreting mammograms of women who are known current users of ERT. Mammographers at GHC receive some clinical information on screened women at the time of interpretation, including ERT use and family history. The knowledge of an association between ERT and increased density and a potential increase in breast cancer incidence associated with long-term ERT use (*36,37*) could raise the preinterpretation level of suspicion and result in more abnormal readings. A lowering of the threshold for declaration of an abnormal reading might also be expected to result in an increase in sensitivity. Such was not the case in our study. Rather, the sensitivity and specificity were lower among current users. Thus, the overall diagnostic accuracy was reduced for current users of ERT.

Sensitivity could be affected by including newly arising cancers in the false-negative group. Some of the false-negative examinations among current users might be explained by the existence of tumors that grew rapidly under stimulation by exogenous estrogen. This acceleration in the growth of spontaneously occurring cancers would lead to an effective shortening of the lead time to diagnosis among current users, bringing tumors to clinical attention sooner in the presence of estrogen than in its absence. This phenomenon is probably less of a factor when one uses a 1-year versus a 2-year interval as the gold standard definition.

Finally, it is possible that the association between ERT use and the decrease in mammographic sensitivity and specificity is the result of some unknown third factor. We were unable, for instance, to adjust for alcohol consumption as a potential confounder (*38,39*), although, for the most part, the known effect of this and other factors on density are modest compared with that of ERT. Consequently, we consider hormonally induced increases in density to be the most likely explanation for the reduction in specificity and sensitivity, especially since this decrement was confined to current users and the effect persisted after adjustment for known confounders. That former users had sensitivities and specificities similar to those of never users suggests that the effect of ERT on screening mammography is transient. Although most current users were taking conjugated equine estrogen (Premarin), there is no reason to suspect that this association is restricted to that particular estrogen preparation.

The impact of even a modest decrease in specificity from 86% to 82% could be significant within the context of a widespread screening program. For example, for a population of 1000 women with a breast cancer prevalence of seven per thousand, ERT would result in an extra 40 false positives. This increase in the false-positive frequency may add considerably to the emotional and economic burdens of screening for breast cancer. At the very least, these false-positive examinations would generate additional mammographic examinations at a reduced interval of 3-6 months (*40*). Some of the false positives would undoubtedly lead to open biopsies, which are considerably more costly from both emotional and economic standpoints. Eddy et al. (*41*) estimated that the cost of a work-up of a false-positive mammogram (abnormal, but ultimately benign, mammographic findings) to be $900 in 1984 dollars. The detrimental psychological costs for the patient and her family are well described (*42-46*). The psychological stress of a false-positive examination is not confined to women with readings that are suspicious for cancer (*42*). If our results are confirmed, these emotional and economic costs will need to be considered in future analyses of the cost-effectiveness of screening programs and of hormonal replacement in postmenopausal women.

ERT has some clear health benefits. The demonstrated reduction in osteoporosis morbidity (*47-49*) and the promise of a reduction in cardiovascular mortality (*50-53*) has made Premarin the most commonly prescribed proprietary medication for the last several years (*54*). As part of the complex, individual decision about whether to use ERT, women should be informed of a possible increase in the chance of a false-positive mammogram while taking estrogen. Future studies should also focus on the duration of ERT's effect on the interpretive process and the benefit of short-term withdrawal of ERT prior to screening mammograms. Although we found a significant decrement in radiographic sensitivity in current users of ERT, this result should be interpreted with caution; the CIs are very wide and, because of the low absolute number of cancer cases, we could not control for factors known to affect the sensitivity of the mammogram. A clinically significant ERT-associated reduction in diagnostic sensitivity could severely compromise the ability of screening mammography programs to reduce mortality. Further studies are needed to investigate this potentially important relationship.

We conclude that the current use of postmenopausal ERT is associated with a reduction in specificity and an increase in the false-positive frequency of screening mammography in women 50 years of age and older. This effect may have important emotional and economic costs. Further study of the relationship between ERT and the sensitivity and specificity of screening mammography is needed, since millions of women and their physicians continue to weigh the risks and the benefits of postmenopausal hormone replacement therapy.

References

(*1*) Fletcher SW, Black W, Harris R, Rimer B, Shapiro S. Report of the International Workshop on Screening for Breast Cancer [*see* comment citation in Medline]. J Natl Cancer Inst 1993;85:1644-56.

(*2*) Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect of screening on breast cancer mortality. J Natl Cancer Inst 1982; 69:349-55.

(*3*) Andersson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Linell F, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. BMJ 1988;297:943-8.

(*4*) Tabar L, Fagerberg CJ, Gad A, Baldetorp L, Holmberg LH, Grontoft O, et al. Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. Lancet 1985;1:829-32.

(*5*) Roberts MM, Alexander FE, Anderson TJ, Chetty V, Donnan PT, Forrest P, et al. Edinburgh trial of screening for breast cancer: mortality at seven years [*see* comment citations in Medline]. Lancet 1990;335:241-6.

(*6*) Frisell J, Eklund L, Hellstrom L, Lidbrink E, Rutqvist LE, Somell A. Randomized study of mammography screening—preliminary report on mortality in the Stockholm trial. Breast Cancer Res Treat 1991;18:49-56.

(*7*) Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Cancer Screening Study: 2. Breast cancer detection and death rates among women aged 50-59 years. Can Med Assoc J 1992;147:1477-1594.

(8) Kerlikowske K, Grady D, Barclay J, Sickles EA, Eaton A, Ernster V. Positive predictive value of screening mammography by age and family history of breast cancer [*see* comment citation in Medline]. JAMA 1993;270: 2444-50.
(9) Fajardo LL, Hillman BJ, Frey C. Correlation between breast parenchymal patterns and mammographers' certainty of diagnosis. Invest Radiol 1988; 23:505-8.
(10) Grove JS, Goodman MJ, Gilbert FI Jr, Mi MP. Factors associated with mammographic pattern. Br J Radiol 1985;58:21-5.
(11) Whitehouse GH, Leinster SJ. The variation of breast parenchymal patterns with age. Br J Radiol 1985;58:315-8.
(12) Anderson TJ. Effects on breast tissue of exogenous oestrogens and progestogens. Acta Obstet Gynecol Scand Suppl 1986;134:9-13.
(13) Saftlas AF, Hoover RN, Brinton LA, Szklo M, Olson DR, Salane M, et al. Mammographic densities and risk of breast cancer [*see* comment citation in Medline]. Cancer 1991;67:2833-8.
(14) Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Do age and density affect breast cancer detection by first screening mammography? J Gen Intern Med 1995;10(suppl):44.
(15) Feig SA, Shaber GS, Patchefsky A, Schwartz GF, Edeiken J, Libshitz HI, et al. Analysis of clinically occult and mammographically occult breast tumors. AJR Am J Roentgenol 1977;128:403-8.
(16) Ma L, Fishell E, Wright B, Hanna W, Allan S, Boyd NF. Case-control study of factors associated with failure to detect breast cancer by mammography [*see* comment citation in Medline]. J Natl Cancer Inst 1992;84:781-5.
(17) Bird RE, Wallace TW, Yankaskas BC. Analysis of cancer missed at screening mammography. Radiology 1992;184:613-7.
(18) Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Cancer Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years [published erratum appears in Can Med Assoc J 1993;148:718] [*see* comment citations in Medline]. Can Med Assoc J 1992;147:1459-76.
(19) Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol Clin North Am 1992;30:187-210.
(20) Peeters PH, Verbeek AL, Hendriks JH, van Bon MJ. Screening for breast cancer in Nijmegen. Report of 6 screening rounds, 1975-1986. Int J Cancer 1989;43:226-30.
(21) Stomper PC, Van Voorhis BJ, Ravnikar VA, Meyer JE. Mammographic changes associated with postmenopausal hormone replacement therapy: a longitudinal study. Radiology 1990;174:487-90.
(22) Kaufman Z, Garstin WI, Hayes R, Michell MJ, Baum M. The mammographic parenchymal patterns of women on hormonal replacement therapy. Clin Radiol 1991;43:389-92.
(23) Berkowitz JE, Gatewood OM, Goldblum LE, Gayler BW. Hormonal replacement therapy: mammographic manifestations. Radiology 1990;174: 199-201.
(24) McNicholas MM, Heneghan JP, Milner MH, Tunney T, Hourihane JB, MacErlaine DP. Pain and increased mammographic density in women receiving hormone replacement therapy: a prospective study. AJR Am J Roentgenol 1994;163:311-5.
(25) Laya MB, Gallagher JC, Schreiman JS, Larson EB, Watson P, Weinstein L. Effect of postmenopausal hormonal replacement therapy on mammographic density and parenchymal pattern. Radiology 1995;196:433-7.
(26) Carter AP, Thompson RS, Bourdeau RV, Andenes J, Mustin H, Straley H. A clinically effective breast cancer screening program can be cost-effective, too. Prev Med 1987;16:19-34.
(27) Taplin SH, Thompson RS, Schnitzer F, Anderman C, Immanuel V. Revisions in the risk-based Breast Cancer Screening Program at Group Health Cooperative [published erratum appears in Cancer 1991;67:2400]. Cancer 1990;66:812-8.
(28) Taplin SH, Barlow W, Urban N, Mandelson MT, Timlin D, Ichikawa L, et al. Stage, age, comorbidity, and direct costs of colon, prostate, and breast cancer care [*see* comment citation in Medline]. J Natl Cancer Inst 1995; 87:417-26.
(29) Sox HC. Probability theory and the interpretation of diagnostic tests. In: Sox HC, editor. Common diagnostic tests: uses and interpretation. 2d ed. Philadelphia: American College of Physicians, 1990.
(30) Breslow NE, Day NE. Statistical methods in cancer research. Vol 1. The analysis of case-control studies. Lyon, France: IARC Sci Publ 32, 1980.
(31) Grove JS, Goodman MJ, Gilbert FI Jr, Mi MP. Factors associated with mammographic pattern. Br J Radiol 1985;58:21-5.
(32) Buchanan JB, Weisburg BF, Sandoz JP, Gray LA Sr, Bland KI. Selected prognostic variables for mammographic parenchymal patterns. Cancer 1981;47:2135-7.
(33) Gravelle IH, Bulstrode JC, Bulbrook RD, Hayward JL, Wang DY. The relationship between radiological patterns of the breast and body weight and height. Br J Radiol 1982;55:23-5.
(34) de Stavola BL, Gravelle IH, Wang DY, Allen DS, Bulbrook RD, Fentiman IS. Relationship of mammographic parenchymal patterns with breast cancer risk factors and risk of breast cancer in a prospective study. Int J Epidemiol 1990;19:247-54.
(35) Jackson VP, Hendrick RE, Feig SA, Kopans DB. Imaging of the radiographically dense breast. Radiology 1993;188:297-301.
(36) Dupont WD, Page DL. Menopausal estrogen replacement therapy and breast cancer [*see* comment citations in Medline]. Arch Intern Med 1991;151:67-72.
(37) Steinberg KK, Thacker SB, Smith SJ, Stroup DF, Zack MM, Flanders WD, et al. A meta-analysis of the effect of estrogen replacement therapy on the risk of breast cancer [published erratum appears in JAMA 1991;266:1362] [*see* comment citations in Medline]. JAMA 1991;265: 1985-90.
(38) Herrinton LJ, Saftlas AF, Stanford JL, Brinton LA, Wolfe JN. Do alcohol intake and mammographic densities interact in regard to the risk of breast cancer? Cancer 1993;71:3029-35.
(39) Funkhouser E, Waterbor JW, Cole P, Rubin E. Mammographic pattern and breast cancer risk among women having elective screening. South Med J 1993;86:177-80.
(40) White E, Urban N, Taylor V. Mammography utilization, public health impact, and cost-effectiveness in the United States. Annu Rev Public Health 1993;14:605-33.
(41) Eddy DM, Hasselblad V, McGivney W, Hendee W. The value of mammography screening in women under age 50 years. JAMA 1988;259: 1512-9.
(42) Lerman C, Trock B, Rimer BK, Boyce A, Jepson C, Engstrom PF. Psychological and behavioral implications of abnormal mammograms. Ann Intern Med 1991;114:657-61.
(43) Baines CJ, To T, Wall C. Women's attitudes to screening after participation in the National Breast Screening Study. A questionnaire survey. Cancer 1990;65:1663-9.
(44) Fentiman IS. Pensive women, painful vigils: consequences of delay in assessment of mammographic abnormalities. Lancet 1988;1:1041-2.
(45) Devitt JE. False alarms of breast cancer [*see* comment citations in Medline]. Lancet 1989;2:1257-8.
(46) Kuni CC. Mammography in the 1990s: a plea for objective doctors and informed patients. Am J Prev Med 1993;9:185-90.
(47) Weiss NS, Ure CL, Ballard JH, Williams AR, Daling JR. Decreased risk of fractures of the hip and lower forearm with postmenopausal use of estrogen. N Engl J Med 1980;303:1195-8.
(48) Ettinger B, Genant HK, Cann CE. Long-term estrogen replacement therapy prevents bone loss and fractures. Ann Intern Med 1985;102:319-24.
(49) Riggs BL, Melton LJ. The prevention and treatment of osteoporosis [published erratum appears in N Engl J Med 1993;328:65] [*see* comment citations in Medline]. N Engl J Med 1992;327:620-7.
(50) Stampfer MJ, Colditz GA, Willett WC, Manson JE, Rosner B, Speizer FE, et al. Postmenopausal estrogen therapy and cardiovascular disease. Ten-year follow-up from the Nurses' Health Study [*see* comment citations in Medline]. N Engl J Med 1991;325: 756-62.
(51) Barrett-Connor E, Bush TL. Estrogen and coronary heart disease in women [*see* comment citation in Medline]. JAMA 1991;265:1861-7.
(52) Lobo RA. Cardiovascular implications of estrogen replacement therapy. Obstet Gynecol 1990;75(4 Suppl):18s-25s.
(53) Psaty BM, Heckbert SR, Atkins D, Siscovick DS, Koepsell TD, Wahl PW, et al. A review of the association of estrogens and progestins with cardiovascular disease in postmenopausal women. Arch Intern Med 1993;153: 1421-7.
(54) Simonsen LL. Top 200 drugs. Pharmacy Times 61:17-23.

Notes

[1]*Editor's note:* SEER is a set of geographically defined, population-based central tumor registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Each registry annually submits its cases to the NCI on a computer tape. These computer tapes are then edited by the NCI and made available for analysis.

Supported by a grant from The Group Health Foundation and by Public Health Service grants PE10002 (National Research Service Award for Research in Primary Care Medicine) and CA63731 (National Cancer Institute) from the National Institutes of Health, Department of Health and Human Services.

We thank Drs. Thomas Koepsell and Bruce Psaty for providing their expert comments and Deborah Timlin for her assistance with data management.

Manuscript received October 10, 1995; revised February 7, 1996; accepted February 28, 1996.